

N O T I C E

THIS DOCUMENT HAS BEEN REPRODUCED FROM
MICROFICHE. ALTHOUGH IT IS RECOGNIZED THAT
CERTAIN PORTIONS ARE ILLEGIBLE, IT IS BEING RELEASED
IN THE INTEREST OF MAKING AVAILABLE AS MUCH
INFORMATION AS POSSIBLE

NASA CONTRACTOR REPORT 166353

E83-10168
CR-166353

"Made available under NASA sponsorship
in the interest of early and wide dis-
semination of Earth Resources Survey
Program information and without liability
for any use made thereof."

ORIGINAL PAGE IS
OF POOR QUALITY

Operational Alternatives for Landsat
in California

Annual Report

(E83-10168) OPERATIONAL ALTERNATIVES FOR
LANDSAT IN CALIFORNIA Annual Report
(Geogroup) 80 p HC A05/MF A01 CSCL 086

N83-17932

Unclas
G3/43 00168

Paul Wilson
M. J. Gialdini

Geogroup

CONTRACT NAS2-11099
July 1981

NASA

NASA CONTRACTOR REPORT 166353

Operational Alternatives for Landsat
in California

Annual Report

Paul Wilson
M. J. Gialdini

Geogroup
2437 Durant Avenue, Suite 206
Berkeley, California 94706

Prepared for
Ames Research Center
under Contract NAS2-11099



National Aeronautics and
Space Administration

Ames Research Center
Moffett Field California 94035

TABLE OF CONTENTS

	<u>Page</u>
ELEMENTS OF VERTICAL DATA INTEGRATION	1
OPERATIONAL ALTERNATIVES FOR LANDSAT IN CALIFORNIA: TECHNICAL ISSUES	41
OPERATIONAL ALTERNATIVES FOR LANDSAT IN CALIFORNIA: INSTITUTIONAL ISSUES	51

ELEMENTS OF VERTICAL DATA INTEGRATION

PAUL M. WILSON
PLANNING CONSULTANT
BERKELEY, CALIFORNIA

ORIGINAL PAGE IS
OF POOR QUALITY

CONTENTS

<u>INTRODUCTION</u>	1
SYSTEMS FOR STUDYING THE ENVIRONMENT	1
ORGANIZATION AND SCOPE OF THE REPORT	5
<u>DEFINING VERTICAL INTEGRATION</u>	7
ESTABLISHING A DEFINITION	7
DATA INTEGRATION V. SYSTEM INTEGRATION	9
<u>ELEMENTS OF VERTICAL DATA INTEGRATION</u>	10
<u>TECHNICAL ELEMENTS</u>	11
1. DATA CLASSIFICATION	12
2. DATA STATUS	13
3. DATA ACCURACY	14
4. DATA VOLUME	15
5. TEMPORAL FACTORS	16
6. GEOGRAPHIC COVERAGE	17
7. DATA REPRESENTATION	18
8. RESOLUTION	19
9. GEOGRAPHIC REFERENCING	20
10. POSITIONAL ACCURACY	21
11. DATA TRANSFER MEDIUM	22
<u>INSTITUTIONAL ELEMENTS</u>	23
1. DIFFERENCES IN NEEDS	24
2. INTERPRETATIONS OF DATA	25
3. LEAD TIME	26
4. USER AWARENESS	27
5. ORGANIZATIONAL RESPONSIBILITY	28
6. COST FACTORS	29
7. STAFFING AND EQUIPMENT	30
8. DOCUMENTATION	31
9. LEGAL BARRIERS	32
10. PUBLIC / PRIVATE ISSUES	33
11. POLITICAL PROBLEMS	34
12. PERSONALITY FACTORS	35
<u>CONCLUSIONS: BUILDING INTEGRATION</u>	36
<u>REFERENCES</u>	39

INTRODUCTION

SYSTEMS FOR STUDYING THE ENVIRONMENT

As man's understanding of the environment has grown, his study of it (and his interactions with it) have become increasingly complex. The number of recognized variables and the interrelationships among these variables has grown at a tremendous rate, putting a large burden on "systems" designed to study the environment.

For an individual or organization interested in the environment of a specific geographic area (whether small or large), an additional problem is encountered: how is the factor of space to be handled? The natural environment being studied will vary in its characteristics from one place to another within the area, and this variation must be taken into account as part of the process of studying it.

The concept of the map has been one response to this problem. A map can be used to portray the spatial variation of resource data over the area of interest. By creating a series of maps at a common scale, an overlay structure can be developed and many types of data for a particular point or area can be compared and analyzed.

When the area of interest is very large or is relatively complex, the requirements of studying that area may tax the traditional methods of using maps. Both portrayal of the data spatially and comparison of different types of data may prove to be too difficult (or too costly or too time-consuming) with manual methods.

As a response to this problem, a number of methods have been developed which use the computer as a tool for the compilation and analysis of resource data. These systems typically include storage of mapped data in a form usable by the computer, a way of performing analytical operations on the data, and a way of translating results into human-readable form.

Many organizations with responsibility for the planning or management of a defined geographic area have developed this kind of system. This is particularly true in California, where areas of local government jurisdiction are often large and complex. A map of existing government-operated systems in the state (Figure 1) reveals the existence of many systems, but with a significant geographic variation. Some parts of the state are covered by several systems (such as a city within a county within a region), while other large areas are not covered at all.

In addition to these governmental systems, there are a number of geobased systems operated by private organizations. Some have been created by forest product companies or by consulting firms for very

RESOURCE INFORMATION SYSTEMS in CALIFORNIA LOCAL GOVERNMENTS



ORIGINAL PAGE IS
OF POOR QUALITY

small areas or for specific applications. Others (particularly those created by the state's two large power utilities) cover large areas and are designed for a broad range of applications. An example of one of these utility systems, operated by Pacific Gas & Electric Company, is shown in Figure 2.

Some parts of the state are apparently not included in any system, while in others there is considerable overlap, with several data bases covering one location. Furthermore, these maps do not include the many state and federal organizations which cover the entire state.

On looking more closely, one discovers that many of these organizations with geobased systems are dealing with the same type of issue, and therefore the same type of data. Most natural resource study depends on such factors as the form of the land, the vegetation and wildlife on it, and the superimposed pattern of human settlement. Typically - almost invariably - each organization collects and stores data about those factors itself.

In addition to the commonality of data types, there is a great deal of spatial overlap. It is easy to identify areas which are within the boundaries of several organizations.

But if several organizations are using the same data for the same area, why is there no sharing of data or systems? That question is the central topic of this report.

The benefits of sharing should be obvious. Total cost is diminished if duplication of data collection and data processing work is eliminated. A greater degree of consistency is insured, so that the decision-making process can concentrate on policy matters rather than questions about data.

Data sharing is obviously good. Its benefits are almost universally recognized. As a concept, it is widely accepted. In practice, examples of data sharing are surprisingly hard to find.

The dilemma, then, revolves around the gap between the potential and the reality of data integration.

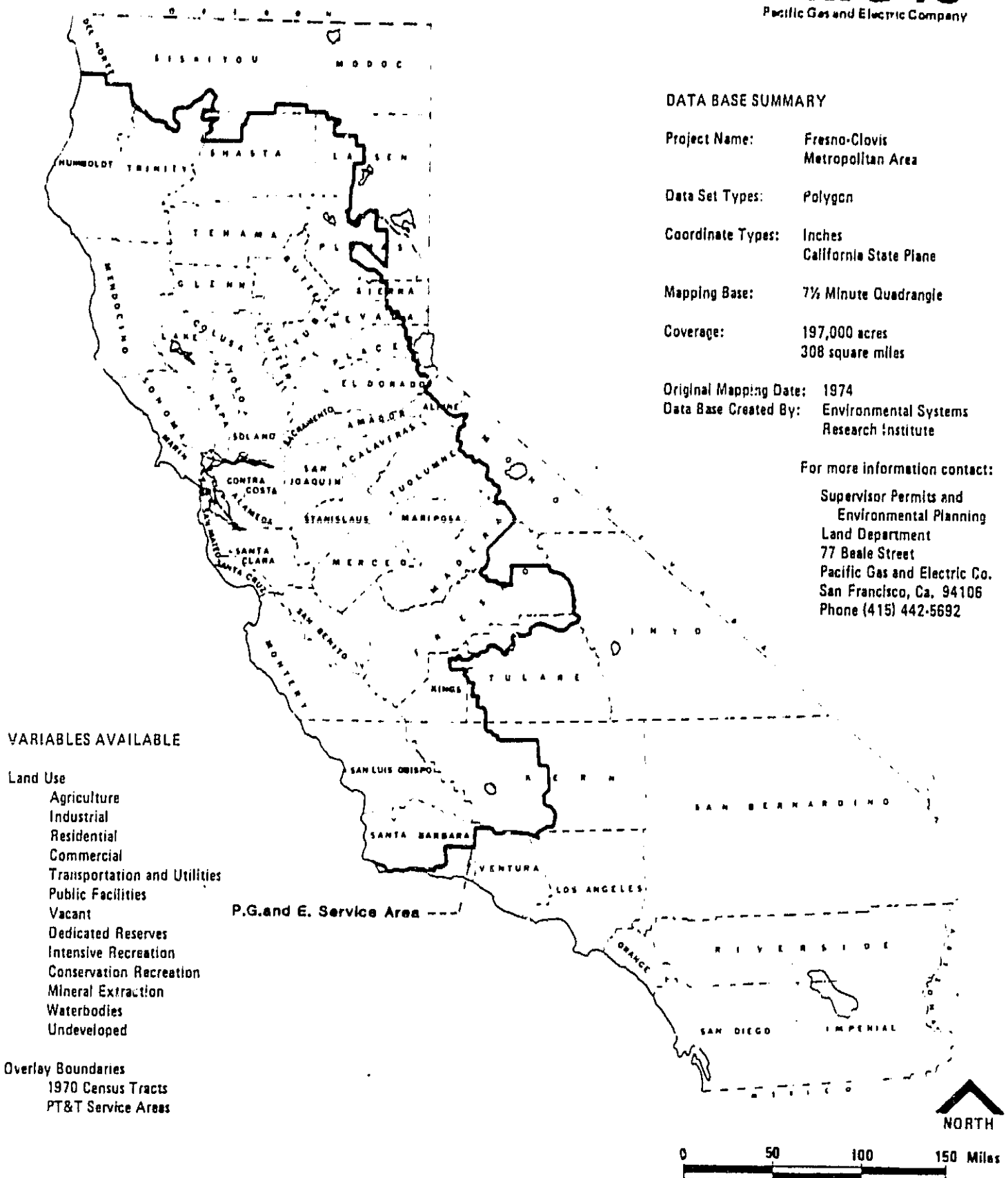


FIGURE 2

ORGANIZATION AND SCOPE OF THE REPORT

ORIGINAL PAGE 13
OF POOR QUALITY

The report is divided into four main parts. In this Introduction, we attempt to create a precise statement of the problem and describe a scope of interest.

The second section focuses on definition of the vertical data integration concept. Since this term is not an accepted part of the field of resource data handling, it is important to reach some level of understanding about its meaning.

Elements of vertical integration are the topic of the third chapter. One major section deals with technical elements (such as resolution and classification) while the other is concerned with institutional factors (organizational control, legal and political barriers).

A final part of the report attempts to fit the theoretical elements of vertical integration into a meaningful structure for looking at the problem from a statewide focus. Included here are general recommendations for achieving a higher degree of integration.

The report is designed to have a scope limited in several ways. The geographic area of concern is the State of California. Although the problems of vertical integration exist everywhere, the current issue is tied closely to potential means of implementation. These are the California Environmental Data Center (EDC) and the California Integrated Remote Sensing System (CIRSS), both of which have a statewide focus.

A second constraint is on the type of data and data systems to be included. The emphasis here is on natural resource data and related land use data. In addition to type, data must also have a geographic reference. These constraints eliminate types such as socioeconomic data (which may have a geographic reference, but do not deal with resources), or environmental data not related to a specific geographic area.

Both manual (mapped) and machine-readable data systems are included, with some emphasis given to automated cases. Special attention is given to LANDSAT imagery as a data source, because of its strong potential for integrated use and its primacy in the CIRSS program.

Finally, the report includes both governmental and private systems. The primary focus is on government, simply because that is where most of the activity (both data supply and data use) has been. Any meaningful attempt to integrate data must, however, include the potential for cooperation between public and private sectors. The federal government is seen primarily as a supplier of data. Lower level governments - state, regional, county, and city - are viewed as both data sources and data users.

ORIGINAL PAGE IS
OF POOR QUALITY

This current report should be reviewed in the perspective of other work, both past and future. A study completed in December 1978 for the California Environmental Data Center, **Resource Information Systems in California Local Government**, is an inventory and evaluation of environmental data base system. (both manual and automated) used by local governments in the state. That survey, along with a broader survey of data systems maintained by the EDC, provides a foundation for assessing the activity (in both data sets and organizations) in the field. These projects help identify candidates for data integration efforts.

Projects currently being conducted under the CIRSS framework look at the ways in which vertical integration can be implemented in an operational setting. Each of the four major projects uses a different approach to integration and is based in a different institutional setting.

The current report fits in between, in both time and subject matter. The emphasis here is on a theoretical structure for thinking about vertical integration. The focus is not on a single case study, but rather on factors which could affect any case. It builds upon the findings of the earlier work, and hopefully will provide useful information for the implementation projects.

ORIGINAL PAGE IS
OF POOR QUALITY

DEFINING VERTICAL INTEGRATION

ESTABLISHING A DEFINITION

What exactly is vertical integration? We have referred to it in a general way above, but it has not been defined with any precision. It is clearly important to do so before proceeding further. Based on the reactions of many people involved in the development and use of resource information systems, it is obviously not a concept which enjoys a universally accepted definition.

We will begin by looking at one draft definition and then continue by examining each part of the phrase separately.

In June of 1979 the California Environmental Data Center recommended the following definition (in a memo from Sally Bay Cornwell):

Vertical data integration (V.D.I.) refers to the general compatibility of data formats, classification methods, and encoding routines, whereby data collected within a geographic area by one agency or level of government can be selectively incorporated into the geobased information systems of many other agencies or levels of government with minimal data manipulation and reformatting.

(Key here are the concepts of integration of data upward or downward between levels of government or across agency boundaries, and the proviso that the data cover a defined geographic locale. Primary emphasis should be given to exploring integration of data between automated geobased information systems, although manual systems should not be ignored, since many local and state agencies have not developed beyond this stage due to the greater initial costs of automated systems.)

Another way of looking at the definition is to analyze the meaning of each word; this procedure should help to precisely delimit what the phrase does and does not mean.

Perhaps the easiest of the words to define is **data**. Here we will treat it as a set of measurements of some resource phenomena intended for some display or analysis. Note that we should preserve some distinction between data and information, as well as between data and its processing. (The next section looks at these distinctions more closely.)

The **integration** part of the term seems to be more clearly understood. This word means a unification, or bringing together of separate parts.

ORIGINAL PAGE IS
OF POOR QUALITY

In the context of resource information, this means use by more than one organization of a particular data set. The data needs of two or more entities have been integrated such that separate collection of the same data is not needed.

Vertical is much more confusing. There are many characteristics of resource systems that can be seen as having a vertical nature. The use of airborne and satellite remote sensing platforms, for example, implies an almost physical concept of the word. Our usage applies to the organizational context in which data is used.

If we assume that there is some integration of data between two agencies, that integration may or may not be vertical. If the agencies are departments of a state or local government, or even adjacent local governments, sharing of data would be horizontal in nature.

The key concept in defining "verticalness" lies in the geographical relationship between the organizations in question. If the two share exactly the same boundaries of interest (as with departments of a government) or have no area in common (as with adjacent cities), the transfer of data between the two would not be vertical. Only when the two entities share some space in common, and have some space not in common, does the possibility of vertical integration exist.

Note that, in a strict sense, a form of vertical integration may take place within a single organization. A parent agency and its division which deals with a subset of the parent's area of concern might also share data in a vertical sense. Our emphasis here is on integration between two or more separate organizations.

Several other points need to be made. First, it should be remembered that data does not have to be used in the same way or the same type of application at each level; integration means sharing the data, and not necessarily its end use.

Also, the difference between compatibility and standardization should be emphasized. Standardization implies a rigid conformance, while compatibility means only the ability to share between users.

Finally, we should note that data integration is seen as a concept. Actual transfer or sharing of data is an embodiment of that concept.

A clear distinction should be made between two levels of data integration, because the operational implications of each can be very different.

Data integration implies only a sharing of data; the data may be in any status. It may be in the form in which it was originally collected, it may be partially reformatted and processed, or it may be part of a larger data base. The thing shared is a unit of data.

System integration, on the other hand, goes a step farther. Here two organizations share not only the data, but the tools for using it. Information processed from the data comes from a common system (whether a computer or a manual map overlay procedure). The "product" integrated, then, may be information rather than data.

There are several characteristics helpful in dividing the two levels. The most useful is physical transfer; if a data set is processed at different places by the organizations, integration would be at the data level only. If each entity uses the same processing system, then the transfer is of information rather than data.

This distinction is often a subtle one; a real example would typically fall somewhere in between the two extremes. The difference is important, however, because strategies for achieving integration are often dependent on what level is to be used.

Total system integration can be viewed as the ideal, since it represents a maximum amount of sharing (and therefore a minimum of duplication). It is critically important, however, to achieve some level of data integration even when full system integration is not possible. Since data collection is usually the most expensive part of building a geobased information system, it is here that the most visible savings can be realized. Data integration can therefore be seen as a logical prerequisite to systems integration.

ELEMENTS OF VERTICAL DATA INTEGRATION

This section covers the detailed aspects of the vertical integration question by identifying elements necessary to achieve integration. Some of these elements are technical: that is, they concern the data base / system itself and the factors necessary to share it. Other elements are non-technical, and are directed more toward the organizational and institutional influences in trying to integrate data or systems.

It is important to keep in mind while reading about the individual elements that they are all closely interrelated. A list such as the one contained in this report could be divided many ways, and several elements could be viewed as different ways of looking at the same problem.

This technical-institutional distinction is, of course, somewhat artificial. Most of the elements discussed below have both technical and institutional aspects, so any division is somewhat arbitrary. It is useful, however, to think about the factors in this way because such a division may suggest the best approaches to achieving compatibility.

ORIGINAL PAGE IS
OF POOR QUALITY

TECHNICAL ELEMENTS

Eleven factors are discussed in this section. They are classed as technical elements because they deal primarily with characteristics of the data itself rather than with organizations which use the data. These characteristics can usually be defined in an objective way. They are, for the most part, measureable. And their status is usually constant (as opposed to some of the institutional factors, which may change rapidly).

The first five elements deal with the how the data is defined and processed. The second grouping of five factors relates to spatial parameters, or how the data relates to geography. The final element covers the technical aspects of data transfer.

1. DATA CLASSIFICATION
2. DATA STATUS
3. DATA ACCURACY
4. DATA VOLUME
5. TEMPORAL FACTORS
6. GEOGRAPHIC COVERAGE
7. DATA REPRESENTATION
8. RESOLUTION
9. GEOGRAPHIC REFERENCING
10. POSITIONAL ACCURACY
11. DATA TRANSFER MEDIUM

1. DATA CLASSIFICATION

Classification is a basic step in any data collection process. It is a decision about the characteristics to be described and the level of detail at which data is to be gathered. Basic questions to be resolved in defining a classification scheme include types and units of measurement (nominal, ordinal, continuous, etc.).

It is important to distinguish between two common usages of the term classification. One is usually applied to an image processing operation used with LANDSAT data. The other meaning is more typical of traditional data collection: a systematic arrangement into categories by a specified set of criteria. That is, of course, exactly what is happening to the raw LANDSAT data.

The issue which makes classification a common barrier to data integration arises from the differences in local information requirements. Each organization which collects data does so to satisfy a specific need, and (in the absence of any motivation to do otherwise) that data will be collected using a classification scheme tailored to that organization's needs. (This difference in needs is an institutional element covered later in this report.)

As a result, each agency adopts its own classification scheme. If that scheme is compatible with the structure used by any other agency, it is more likely to be the result of luck than design. When data is transferred, it will be necessary to convert it so that it fits into the classification structure used by the second organization. Sometimes such a correspondence cannot be made, and the choice becomes one of either not using the data or changing the current classification structure.

Other characteristics of a classification scheme may arise from the limitations of the system being used. A map may become too complicated if the categories are too numerous. In a computer-based system, a coding structure may be limited by the number of bits available to code each data item.

Where there are widely-accepted (or unique) sources for one particular type of data, the original categories will suggest a classification scheme. Soil and geology maps are examples of this. The availability of some commonly-accepted structure such as the USGS land use definitions is helping to alleviate the problem. Adoption of other data sources (e.g., LANDSAT imagery) to standard classifications is a major contribution to vertical integration.

2. DATA STATUS

The distinction between data and information is a key factor in using resource (or any other) data. This distinction becomes important in the vertical integration context because there may be compatibility between the data needs of two organizations but no compatibility in information needs.

Data may be collected or stored in different levels of thematic aggregation. A data set may reflect grain size, soil group, or prime agricultural soil designation, according to the needs of the agency collecting it. Similarly, a data set containing elevations may be converted to one showing slope or aspect.

If the process used by the data source agency to derive information from the base data destroys that data (or if, for whatever reason, the original data is not retained), the possibility of data integration is lost. Transfer is then possible only if the information needs of the organizations are identical.

The solution to this problem is obvious: the data should be retained in its original form whenever possible. This is usually feasible if the only cost is for the recording media (maps or magnetic tapes). A further problem is presented if maintenance of the data (correction and update) is done only on some processed version and is not applied to the original data set.

3. DATA ACCURACY

The element of data accuracy pertains to the thematical quality of the data. It is the degree of error in identifying and measuring the characteristic under observation. The important issue here is not the positional accuracy of an individual observation - that is the subject of another element discussed below - but the absence of error in assigning a particular point to a classification category. For example, it is important that in a land use data file an area of residential use be assigned to an appropriate category. Where data cannot be classified properly, it may be better for it to be assigned to an unclassified or unknown group rather than to be put in an incorrect category.

Several aspects of data accuracy can be considered. Absolute accuracy is difficult to attain even in an ideal setting, and virtually impossible in any operational environment. Improving the level of accuracy beyond some acceptable point is likely to become increasingly expensive as total accuracy is approached. There should be a distinct relationship between accuracy and costs: attaining a higher degree of accuracy will inevitably cost more and increase time requirements.

A final, but vitally important factor: accuracy requirements will vary widely for different users. Continuing with the land use example, a statewide survey may need only to identify land under a single broad residential category, while a local government would probably require division into several categories by density or type. (The classification issue again.) The locally-produced data set may have many cases of misclassification between residential categories (e.g., an apartment buliding erroneously marked as single family residential) but would be perfectly accurate for the purposes desired by the state.

ORIGINAL PAGE IS
OF POOR QUALITY

4. DATA VOLUME

Volume is an easily understood factor, and is fairly easy to calculate. It is simply a measure of the amount of data which must be transferred for integration to take place.

Volume is a direct function of two of the other factors, resolution (number 8) and frequency (a part of number 5). High resolution implies a larger volume of data for a given area. Similarly, more frequent collection of data means higher volume also. (This assumes that all of the data collected is to be transferred. If there is no need to use every data set, then volume might not become a problem for the user agency.)

LANDSAT is a prime example of volume as a potential barrier to data integration. It "suffers" from both high resolution and a frequency that is extremely high in comparison to other sources of natural resource data. This volume is, of course, one of the characteristics which makes LANDSAT a good data source. At the same time, such a large volume can severely restrain attempts to integrate data because the channels of data transfer might be unable to cope. Local users, who typically do not have in-house capability to process LANDSAT data, are particularly susceptible; they can be easily overwhelmed by the volume.

Solutions to the volume barrier would seem to lie in reducing either resolution or frequency. Aggregation of high resolution (e.g., pixel) data into larger grid cells or polygons would reduce the number of data elements in any single data set. Reduction of frequency could be accomplished by selecting only those data sets which were of most interest.

A third technique seems obvious, but is often ignored because of technical reasons. The volume of data to be transferred is often unnecessarily high because the data set covers an area larger than the one of interest. This is particularly true of LANDSAT, where data sets are normally dealt with in terms of scenes. A user agency must extract the data concerning its area of interest from the collection of scenes which contain it; a geographic "filter" would cut data volume substantially in most cases.

5. TEMPORAL FACTORS

A number of factors can be grouped together by a common thread of time. They concern the frequency with which data is collected, its timeliness, the time required for processing, and historical integrity.

Frequency is perhaps the most obvious temporal factor. Applications which rely on the assessment of change require a data source which is generated or updated fairly frequently. LANDSAT data has a very high frequency, especially in comparison to most of the conventional data sources with which it might compete.

Of equal concern is timeliness. Data collection must be timed to reflect desired characteristics. The timing may be dependent on single natural events (e.g., the extent of a flooded area), a human-scheduled event (such as a land use survey timed to coincide with the collection of census data), or seasonal variations (e.g., a crop inventory or vegetation surveys). Clearly, if the timing of data collection is not compatible with the needs of the user agency, data integration will not happen.

The time required to process data is another important factor. A data set collected in a timely manner is of no use if it cannot be processed and transferred before it is needed. This, too, has been a limitation on the use of LANDSAT data; although frequency of data collection is high and timeliness is usually acceptable, the long lead time required to acquire and process imagery often prevents its use.

Historical integrity is a more subtle, but still important, temporal factor. Many applications require the analysis of data collected at different times (perhaps several years apart), and this data must be comparable across the other technical elements (classification, accuracy, etc.).

6. GEOGRAPHIC COVERAGE

The area covered by a data set is a primary determinant of its usability by several organizations. Each agency which collects data does so only for its own jurisdiction or area of interest (unless somehow motivated beforehand to extend the coverage). The reason for this mode of behavior is obvious: dealing with a larger area will almost always cost more money.

Chances for vertical integration depend largely on the spatial relationship between two areas: one, the jurisdiction of the data collecting agency, defines the area for which data will be available; the second is the jurisdiction of the potential user, and defines the area for which data is needed. (Note here that the word jurisdiction is used to refer to an area of interest for a specific project, and is not necessarily a legally-defined area under the scope of a governmental agency. It may be a subset of a governmental boundary, or a project area of interest to a private firm. The area of interest may vary widely even within the same organization, as different applications focus on different areas.)

If the data coverage is larger than the area of interest, then the problem is one of extracting the needed window from the larger data set. This is commonly the procedure used with LANDSAT imagery or most federal map sources. If, on the other hand, the area for which data is needed is not completely contained within the area bounded by the available data set, other sources must be found to fill in gaps in the coverage.

Obviously, the first situation is easier to deal with than the second. That is one reason why federal data sources are often attractive for state and local governments, while data collected at a statewide level is frequently used by local governments.

ORIGINAL PAGE IS
OF POOR QUALITY

7. DATA REPRESENTATION

Data representation refers to the spatial mode in which the data is collected or stored. In manual systems for handling natural resource data, this mode is usually a map. In an automated GIS, grid cell and polygon structures are most frequently used. In some cases, data is aggregated by a named or numbered administrative unit.

Problems occur when the potential user of a data set bases its system on a mode of data representation different from that of the data supplier. The issue then becomes one of transformation between modes.

This kind of problem can often be overcome using automated methods to change modes. Conversion of polygons to grid cells is a common procedure, and usually produces satisfactory results. The reverse shift - grid cells to polygons - is a different matter. Some amount of accuracy is lost, depending on the size of grid cells being converted.

The worst problems occur, however, when there is no specific X-Y coding of data at all. Aggregation of data by some sort of administrative unit (such as a city boundary or census tract) loses the exact representation of data within that unit.

This question of aggregation leads to one of the more critical aspects of data representation: the distinction between the mode in which data is collected and the mode in which it is stored. For example, data stored in a grid cell data base often originates in a polygon format (or, if not a true polygon, some digitized structure such as vectors or chains which could be assembled into polygons). If an agency desiring to share that data uses a polygon representation for its processing, problems are minimized if the data in its original form is available.

8. RESOLUTION

Resolution is the discriminating power of the spatial units used for data representation. It can be viewed as the smallest unit of data storage in "ground" terms. In transfer of data between manual systems, resolution is closely related to map scale. With digital data, resolution can refer to the smallest physical unit of data collected or it can mean the units in which data is stored.

There are, of course, practical limits on the fineness with which data can be pinpointed. Most surveys use a minimum size unit for data collection (such as a one-acre minimum for a land use file). Other modes of data collection are limited to a given resolution for technical reasons; the pixel format of LANDSAT imagery is a good example of this.

Storage resolution is usually a matter of choice. Data cannot, of course, be stored at a resolution finer than that at which it was collected; it is often aggregated to a standard cell size for insertion into a data base.

The barrier to data integration presented by the resolution factor is well known. If the user agency needs a finer resolution than that at which the data is collected, there will probably be no transfer. The question of resolution is frequently mentioned as a major limitation to the use of LANDSAT data in some types of applications. This barrier may be overcome to some extent by the smaller pixel size projected for future satellites.

The LANDSAT example points to a general dilemma in vertical data integration: agencies which collect data for large geographic areas (eliminating factor 6) usually are limited in the resolution with which they can cover those areas. There is a critical balance between the two goals of coverage and resolution.

9. GEOGRAPHIC REFERENCING

ORIGINAL PAGE IS
OF POOR QUALITY

Representing data from a globe in a two-dimensional form usable by humans involves the creation of a projection and an accompanying coordinate system. The geographic referencing element deals with the form of representation used.

Maps are constructed around a particular projection (although they may contain tic marks representing several coordinate systems). Manual transfer of data is inhibited if mapped data sources use a different coordinate system than does the potential user.

In automated systems, mathematical transformation between standard coordinate systems is usually straightforward. Problems occur only where local or other nonstandard forms are used. As long as standard types (latitude-longitude, UTM, State Plane) are used the solution depends only on the availability of appropriate software.

10. POSITIONAL ACCURACY

ORIGINAL PAGE 13
OF POOR QUALITY

Another important technical element is positional accuracy, or the absence of error in applying the geographic referencing system to the actual data. Problems can occur in several places during the data collection process that contribute to positional errors. These may take place during original data collection, in changing mode (such as converting a digitized file to polygons or grid cells), or in resampling (pixels to cells).

As with many of the factors already discussed, the question of positional accuracy has varying degrees of importance for different users and different applications. A given amount of error (say a shift of 10 meters from a true position) may be very important in a local government land use survey, but could be ignored in a statewide project.

The requirements for positional accuracy also vary with type of data. Anything that deals with property lines, for example, or is important for an engineering application, should have a high level of accuracy. Data that is more ephemeral, such as air pollution readings, need not be as accurate.

11. DATA TRANSFER MEDIUM

If all of the above technical elements have been satisfied and the decision has been made to proceed with sharing data, the question about the actual means of a transfer is encountered.

Maps will be the primary medium used to transfer data between manual systems (or between a manual and an automated system); they may be accompanied by statistical tables or other supplementary information. Maps are also a possible mode of transfer between two automated systems where transmission of digital data is not feasible.

Transfer of data by map is made somewhat easier by the prevalence of common base maps. Many local government base maps are derived from USGS sources, and therefore will show at least a minimum level of compatibility. The adherence to common scales also is helpful.

Ideally, data passing between two automated GIS users should remain in machine-readable form to minimize conversion costs at both ends. Transfer of digital data requires close attention to factors such as coding conventions, file structures, tape density and formats, and transmission speeds.

The growth in telecommunications capabilities raises the possibility of networks where data could be retained by the original source agency (or perhaps at central data storage centers) and transmitted to users when needed.

INSTITUTIONAL ELEMENTS

A second set of concerns deals not with the data itself, but with the organization environment in which the data must be transferred. These elements are qualitatively different from the technical ones; they are harder to define as concepts, harder to identify in a real setting, and much more difficult to overcome if they become barriers to data integration.

1. DIFFERENCES IN NEEDS
2. INTERPRETATIONS OF DATA
3. LEAD TIME
4. USER AWARENESS
5. ORGANIZATIONAL RESPONSIBILITY
6. COST FACTORS
7. STAFFING AND EQUIPMENT
8. DOCUMENTATION
9. LEGAL BARRIERS
10. PUBLIC V. PRIVATE ISSUES
11. POLITICAL PROBLEMS
12. PERSONALITY FACTORS

1. DIFFERENCES IN NEEDS

Much of the difficulty in making institutional connections is based on one simple and obvious premise: different types of organizations have different data needs. Whether these needs are derived directly from laws under which the organization operates, or whether they have been defined internally for technical reasons, they are very difficult to change.

It should be emphasized that all of these differences are not bad. Each organization must respond to the pressures of its own environment; many differences are natural, and it would be a mistake to force conformance when it is not called for. Some of the differences, however, are more artificial. They may result from a perceived need which has no valid connection to the organization's work; the original program which called for a certain type of data may have substantially changed.

These examples of resistance to change form a real inertia around accepted ways of collecting, storing, and manipulating data. "We have always done it this way" is all too often the end to a data integration proposal.

Much of this resistance can be traced to a lack of a common conceptual framework in many fields. The emerging disciplines of planning and environmental protection have yet to settle on ways to attack certain problems, so similar agencies may have very dissimilar approaches to the same problem. The result is a lack of compatibility in data needs.

2. INTERPRETATIONS OF DATA

Observations of the world are often colored by organizational factors. The same thing seen from the perspective of two different agencies is likely to be recorded differently unless there is some common framework for interpretation.

Ideally, every data collection project would utilize a totally objective approach. Distortions from "reality" would be minimized. Making this concept operational is extremely difficult; defining reality as it applies to resource data is a considerable problem in itself.

When any data is brought into an organization from outside, it must be reconciled with internal sources. Any unacceptable interpretations added by the original collector of the data must be removed. This process can be extremely demanding, and could add considerably to the cost and time required for data transfer. It therefore can act as a major barrier.

3. LEAD TIME

Time is an important factor in using data. Several technical elements related to time were discussed in a previous section. There are other aspects of time which are more organizational in nature; these relate to lead time and deadlines.

Many timing problems, like the differences in needs discussed above, are very real. An organization which must respond to a forest fire hazard has very apparent needs for data that can be obtained quickly. At the other extreme, a planning agency building a land use file to coincide with census surveys has a very long period in which to schedule its work.

Internal deadlines can often force duplication of data because the project schedule does not allow time to transfer existing data. The root problem is typically a lack of long range planning for data needs. It is all too common for the time requirements of data collection, editing, and reformatting to be ignored.

Another common misconception is the difference in time between the point at which data is collected and the point at which it becomes usable. The many necessary components of the process which come after collection are forgotten.

A final dilemma is the unanswerable question of perfect data v. available data.

4. USER AWARENESS

Lack of knowledge about the existence of a needed data file is one of the simplest reasons for failure to integrate. In this age of massive data banks and telecommunications networks, it seems somehow astounding that this factor remains such an important barrier. In many cases, however, needless duplication results simply because a user was not aware of the existence or availability of a certain data set.

If the user organization has no desire to use existing data, there will, of course, be no transfer. This may point to other institutional factors as problem areas. The concern here is with the organization that would like to obtain data that already exists rather than re-collect it; but simply does not know where to look.

For most users of resource data, there is no regular means of communicating about data sources. This is a serious problem even within one level of government (such as between cities in California); it is even more prevalent between levels. Vertical integration is made even more difficult.

Some progress has been made in this area. In its relatively short span of existence, the California Environmental Data Center has become a clearinghouse for data exchange within the state. At the national level, several large data collection agencies (NCIC, EROS) provide directories to existing sources.

5. ORGANIZATIONAL RESPONSIBILITY

Data does not transfer itself. Some organization or individual must assume the responsibility for making the connection between provider and user. That connection may involve several stages - the data must be collected, edited, perhaps updated, reformatted, stored, and analyzed - and there may be a separate organization in charge of each step. The important point to be made is that some organization must take that responsibility; integration will not happen by itself.

This element frequently becomes a barrier to data integration because there is usually no clear assignment of responsibility among organizations. Unless there is a broker, whose function it is to facilitate the transfer, most of the burden of making the connection falls upon either the data supplier or the user agency.

An underlying problem is the question of incentives. The potential user has a basic incentive for attempting to integrate - the availability of a desired data set. If there are major barriers to be overcome (significant incompatibility in either technical or institutional factors), the incentive to share may be overwhelmed by the perceived costs. A common result is to forget the potential integration in favor of a separate data collection effort.

The incentive issue can be even more burdensome from the standpoint of the organization supplying the data. What benefits does the supplier get for its efforts? For a public agency, the only expected benefit may be a recovery of costs - either the cost of developing the data set or only the actual cost of transfer. A private firm may reasonably expect a greater return on its investment when it supplies a data set. When sufficient incentives are absent, it will be very difficult to motivate the holder of a data set to transfer it.

One possible answer to this problem is the strengthening of the broker role. Both NASA and EDC provide some of these services in the CIRSS context, and their efforts help overcome many potential barriers. Not to be overlooked is the role that the private sector can play in making connections.

6. COST FACTORS

ORIGINAL PAGE 13
OF POOR QUALITY

Cost is a prominent consideration in acquiring and using data. If the costs (actual or perceived) of transferring data are too high, no data integration will take place.

The important aspect of cost from the user standpoint is not the nominal cost of raw data, but the total cost of putting that data in a usable form. Any of the technical and institutional factors can raise this total cost. For example, extensive reformatting or geometric transformation of a data set may be many times more expensive than the original data itself. LANDSAT is a perfect example of this: the initial imagery is very inexpensive in comparison to other sources for similar data, but processing costs can be prohibitively high.

Even when the total costs of acquiring a data set are recognized, there may be a further barrier in insuring a sufficient budget. Resource data often has a high front end cost in comparison to operating costs. In many public agencies, budgets for data handling may be so fragmented among departments or programs that it is impossible to collect enough to match total cost.

Continuity over time is a final aspect of cost. Most resource data is useful only in the context of a long-range program; a system to acquire and use this type of data must be supported over a period of several years to be feasible. Given the nature of the public budgeting process, such continuity is not easy to achieve. Cost limitation efforts like California's Proposition 13 make it even more difficult to budget for long range projects.

7. STAFFING AND EQUIPMENT

Processing resource data is often highly technical work, requiring special equipment and staff skills not found in many organizations. Being confronted with the necessity to augment in-house capability to be able to use a certain data set can be a major barrier to integration.

LANDSAT suffers frequently from this barrier. The extensive image processing work required to use LANDSAT data is beyond the current capabilities of most user organizations. This is especially true in government agencies, where the cost factors noted above prevent acquisition of necessary staff and equipment.

Answers to this problem do exist. Help is frequently available from federal or state government agencies; NASA's role in the current CIRSS projects is a good example of this. Also, assistance can be obtained at some cost from private sector firms or university programs which specialize in this type of work.

8. DOCUMENTATION

Inadequate records of data collection or analysis methods can be a barrier to reuse of data sets. Poor documentation can greatly increase the cost of transfer if previously-completed work must be duplicated. Furthermore, a lack of documentation can raise uncertainties about the applicability of a given data set; its potential for integration will be much harder to evaluate.

Several types of documentation can be identified. Probably the most critical is that dealing with the technical factors covered earlier. Successful transfer of data depends on achieving compatibility among all these elements, and the task is made more difficult if there is no known set of data characteristics from which to work.

Another level of documentation is more applications oriented than the technical type. This kind of documentation helps to publicize the existence of data sets, and thereby contributes to solving the user awareness problem.

One real gap in documentation has been in the attention to institutional factors. Typical project documentation concentrates on technical findings and ignores the setting in which the project took place. This tendency unfortunately prevents learning from the organizational experiences of others.

9. LEGAL BARRIERS

ORIGINAL PAGE IS
OF POOR QUALITY

Transfer of data is sometimes prohibited (or at least inhibited) by laws. This can be true both for public agency data (such as census surveys) and privately-held data (such as mineral surveys which map contain proprietary information).

Some of these legal considerations have real value. Privacy is a major concern with some types of data, although it usually does not apply to the resource type that is our focus here. Protection of the investment symbolized by a mineral survey is important too.

Some types of resource data do require care in distribution. Any map or data set which identifies the location of fragile features (archeological sites, endangered wildlife habitats) should be transferred only if there is assurance of proper care from potential users.

Liability is a legal factor that is often ignored in integrating resource data. What legal responsibility does the data supplier have for the accuracy of the data set? Does this responsibility extend to applications of the data by others? This type of question can result in a barrier to transfer by increasing the reluctance of data source organizations to share.

10. PUBLIC / PRIVATE ISSUES

There is a frequent unwillingness to share data between public and private organizations. Such a barrier can prevent integration of data even when all of the other elements have been satisfied.

Cost is often cited as the reason for this problem. Public agencies may be reluctant to pay for data collected by a private firm, or - if they have a data set needed for some private sector application - be unwilling to release "their" data. Cost (which is discussed more fully as element 6 in this section) is a factor, but the problem goes deeper than that.

The underlying dilemma is a feeling of mutual distrust in many public-private relationships. This may spring from situations where the two sides are in adversary positions; resource data is frequently used as a tool in debates over project proposals (as in the environmental impact review process) or governmental controls (air pollution mitigation programs). Sharing of data may be perceived as sharing ammunition with the enemy.

Perhaps the only way to minimize this factor is to point out the mutual advantages of public-private cooperation. Formal channels of communication between the two sectors (such as the CIRSS Task Force and Industry Advisory Panel) are helpful in this regard.

11. POLITICAL PROBLEMS

It is often said that information is power. Shifts in who holds information, then, can lead to shifts in power. Such shifts will inevitably be seen as undesirable to the current holders of power and beneficial from the standpoint of those organizations wanting more.

This strong desire to maintain power (which can be viewed as organizational self preservation) can be a strong barrier to data integration. An organization controlling a particular data set will be very reluctant to make that data available to any other organization which it perceives as a potential threat.

Environmental data has become more and more susceptible to this sort of political manipulation as its importance in the decision-making process has grown. Conflicts between levels of government over some resource issue are a direct barrier to vertical data integration. Each side may feel that the other's data has been slanted to support its own position; the result is a more or less deliberate duplication of what should be an identical data base.

Another political factor is public perception. If there is a fear that the public will disapprove of expenditures for some technical project (such as a large data collection effort), then an agency will be reluctant to publicize it by making it available to others. This is particularly true if the nature of the project is controversial; a data base containing feasible sites for an unpopular activity (such as a nuclear plant) might be closely protected.

An underlying problem with gaining public acceptance is the difficulty in measuring the benefits of a GIS. Most benefits are long range; they lie in optimizing the future use (or avoiding the misuse) of natural resources). This factor makes it even more difficult to build a user community, both inside and outside the organization.

12. PERSONALITY FACTORS

The previous element dealt with perceptions of power and fear at the organizational level. This section looks at the same sort of problems at the level of individuals within the organization. The essential difference is one of scale.

These factors, like political problems, are largely intangible. They may result from very different types of motivation. Some are based on fear: suspicion of the reasons for wanting data, loss of control over vital information, or simply a general distrust of new technology. (The latter feeling is often expressed about LANDSAT, which represents a major departure from the way resource data has traditionally been collected and handled.) Other motivations spring from a desire to increase individual power or status; empire-building around an information base is all too common.

Another problem area can be the basic incompetence of an individual who must play a key part in the data transfer process.

Personality factors are often the most difficult problems to deal with when they get in the way of data integration. They are hard to assess in an objective way, and often impossible to change even when they can be identified.

CONCLUSIONS: BUILDING INTEGRATION

Thus far we have defined vertical data integration and identified the elements which may act as barriers to achieving it. Here the topic is solutions: how each of these factors should be addressed so that it does not prevent integration in an operational setting. Given the diversity of data uses and users, there can be no real rules; the ideas here should be treated as general guidelines which must be modified to suit particular situations.

An important initial consideration is the requirement that all of the elements discussed in the previous section be satisfied for successful transfer of data. Data integration is fragile. It is not sufficient to meet some of, or even a majority of, the requirements; the failure of any one element can prevent data integration.

Often one technical element, if not compatible between a data source and a potential user, can stop transfer. With LANDSAT data, for example, resolution is often noted as a primary reason for infrequent use in local government applications. Even though LANDSAT provides a source of data which exceeds requirements of any user in most categories (geographic coverage, temporal factors, etc.) an inadequate level of resolution can be enough to outweigh the many positive factors.

Similarly, a single institutional factor can also prevent integration. This is often the case where the data may be adequate in all technical elements, but legal or political barriers get in the way of vertical reuse. More commonly, the organization in need of data simply may not know about the existence of a data set which could satisfy its requirements.

Institutional factors such as cost are barriers where the data source is in the private sector and expects a profit from use by others. This cost element is not directly important with LANDSAT, since tape and photo products themselves are not expensive. Processing of these products, however, can have significant costs.

What happens when transfer of data is not possible because one or more of the elements is not satisfied? From the standpoint of the potential user - the organization with a defined need for a data set and unable to meet that need through vertical transfer - there are three broad types of response. One is to accept the unavailability of the necessary data by cancelling the need; the application is ended or redefined to eliminate the need for that type of data. A second response is to

ORIGINAL PAGE IS
OF POOR QUALITY

continue the project, but to find another source for that data. This solution, of course, is likely to have heavy time and money costs.

Neither of these responses includes vertical integration. A third approach, which does support the vertical data integration concept, is to overcome the barrier to transfer; integration is achieved by eliminating the factor or factors which prevented it. The elimination of these barriers will not, of course, be without costs: time and money will be required.

Realistically, any attempt to transfer data between agencies will meet some difficulties. Even if there is some degree of compatibility in all factors and transfer is possible, it would be highly unusual for there to be an optimal fit in every category. The question becomes one of gaining an acceptable degree of compatibility between the available data set and the data required for the application at hand.

Timing is a key issue. It is better to insure compatibility early in the planning process. Key technical decisions are made very early in any data collection project. If these decisions are made in a way that limits compatibility, integration may never happen.

Ideally, barriers should be removed in the data definition stage rather than having to worry about making existing data fit an existing need. The costs of achieving compatibility increase as the data collection and processing steps are completed; the best approach is to insure compatibility early by making integration an important consideration in the planning process.

In general, it appears that the technical issues are easier to solve than the institutional ones. It is much harder to even identify organizational problems because they are usually specific to the setting. A given technical problem can usually be solved by a method that has worked before. This is not necessarily true for an institutional issue.

It is also true that there is less attention paid to the nontechnical problems. Less research has been done on them, and the organizational findings of successful projects are not communicated very well.

Perhaps the key problem, of all the elements covered in this report, is that of organizational responsibility. In most cases there exists no recognized broker, no entity responsible for facilitating integration by making the necessary connections between source and user. Individual agencies rarely have sufficient motivation or skill to make the connections by themselves.

The Environmental Data Center is perhaps the only existing agency in California with the geographic scope and setting necessary to serve in this role. To be more effective, it might soon be advisable to move from a user awareness role to one that includes the actual transfer of data.

ORIGINAL PAGE IS
OF POOR QUALITY

This report was designed to serve as a first step in building a process for integrating resource data in California. It has attempted to raise the consciousness of those working in this field by identifying the elements which are important in achieving integration.

The current operational projects funded under the CIRSS program will take the process one step further. Each of the four projects is designed to look at integration in a different institutional environment. Each can be viewed as one major option for building vertical integration in the state.

The concepts discussed in this report will be tested in those projects. Recommendations for specific actions (by NASA, EDC, the Task Force, other governmental agencies, the academic community, and the private sector) will be made. A strong foundation for vertical integration should result.

REFERENCES

Discussion of both technical and institutional factors important to geobased information systems is contained in several sources.

Center for Geographic Analysis, University of Wisconsin. Data Needs and Data Gathering for Areas of Critical Environmental Concern. (Madison, 1975.)

Natural Resource and Environment Task Force of the Intergovernmental Science, Engineering, and Technology Advisory Panel. State and Local Government Perspectives on a Landsat Information System. (Washington: Executive Office of the President, 1978.)

Michael Kennedy and Charles Guinn. "Avoiding System Failure: Approaches to Integrity and Utility," in Technical Supporting Report E of Information / Data Handling: A Guidebook for Development of State Programs. (Washington: Department of Interior, 1975.)

William Toner, Elizabeth Hollander, and Judith Getzels. "Stonewalled: Barriers to Adoption of Computer Oriented Land Use Information Systems," in Technical Supporting Report E of Information / Data Handling: A Guidebook for Development of State Programs. (Washington: Department of Interior, 1975.)

R.F. Tomlinson, ed. Geographical Data Handling. (International Geographical Commission on Geographical Data Sensing and Processing, 1972.)

Sources of information about the California situation include publications of the Environmental Data Center and the CIRSS Task Force.

Paul M. Wilson. Resource Information Systems in California Local Government. (Sacramento: Office of Planning and Research, 1978.)

OPERATIONAL ALTERNATIVES FOR LANDSAT IN CALIFORNIA:

TECHNICAL ISSUES

M. J. Gialdini

ESL Incorporated

OPERATIONAL ALTERNATIVES FOR LANDSAT IN CALIFORNIA:
TECHNICAL ISSUES

The California Integrated Remote Sensing System (CIRSS) is a NASA-funded project aimed at investigating the concept of vertical data integration within California to include the potential of Landsat for meeting data needs within the state. In 1980, CIRSS contained four demonstration projects designed to examine four different approaches to vertical data integration: evolutionary, network, developmental and industry transfer. A fifth project, the operational alternatives study, was directed at evaluating the efficacy of those four approaches in achieving vertical data integration. This evaluation focused on both the technical and institutional issues encountered within the implementations of those approaches and not on the specific successes of the projects themselves.

As a concept, vertical data integration, unfortunately, is very difficult to define. After lengthy discussion, the CIRSS Task Force established the following definition as a guideline:

Vertical data integration refers to the general compatibility of data formats, classification methods, and encoding routines whereby data collected within a geographical area by one level of government and its associated agencies can be selectively incorporated into geobased information systems of many other levels of government and their associated agencies with minimal data manipulation or reformatting.

An excellent discussion of the conceptual issues surrounding vertical data integration is in a report to the CIRSS Task Force titled Elements of Vertical Data Integration by Paul Wilson, 1980. This work formed the framework for the evaluations reported both here, covering only technical issues and in a companion report, Operational Alternatives for Landsat in California: Institutional Issues,

ORIGINAL PAGE IS
OF POOR QUALITY

again by Paul Wilson, May 1981. Reading the "Institutional Issues" report is strongly recommended as it contains valuable background information concerning CIRSS, the Task Force and the Operational Alternatives Study that, in the interest of avoiding unnecessary redundancy, will not be included here.

The remainder of this report will cover the evaluations of the four approaches focusing on technical issues and is based on information drawn from numerous interviews with project participants: performers and users.

Evolutionary Approach

This approach was intended as an extension of the work that produced the California Department of Forestry (CDF) statewide Landsat classification. This was to make that data available to other parts within CDF and county governments and see how it could evolve into other uses or be combined with other data sets to satisfy other data needs. A basic assumption in this approach is that the CDF capabilities would evolve to a point where it could support local agency vertical integration projects.

One of the most significant results to date is the creation of a GIS demonstration for Santa Cruz County that had participation by local, state and federal agencies. The objective was to demonstrate the cooperative development of a common data base and to exercise it for the needs of each contributor. The method was to add layers to existing Santa Cruz geographic data to bring it to a condition indicative of an operational data base and then produce operational and/or experimental demonstration results with the participation of each user.

CDF has also been working to define the hardware and software capabilities needed to create an operational system for use in the state offices. This system would, in effect, initially be a replacement for the support provided by NASA but also provide the basis for CDF to further develop their in-house operational capabilities.

ORIGINAL PAGE IS
OF POOR QUALITY

Technically, this approach enjoyed a situation of having prior experience with Landsat data processing on CDF staff by virtue of the work in developing the statewide classification with NASA. As a result, there was little difficulty in working with that particular data set. The GIS demonstration in Santa Cruz County was a good example of vertical data integration concepts; however, there is some question as to its leading to an operational system. This is primarily due to the heavy emphasis on the use of multiple computer systems employed by NASA technical representatives in performing the work. These systems represent a significant capital investment, perhaps beyond the capacities of the user government agencies involved; and requires skilled and experienced analysts to properly apply the power of these resources to a given problem. While it is conceivable that a single system could be applied--the analyst's technical background requirement would still be substantial. The fact that CDF is looking into hardware and software to build their own system is a positive step but they must be willing to accept the time and expense required to train the staff at all levels. The combination of an operational system and skilled staff would give CDF an excellent opportunity to explore vertical data integration further. From a technical point of view, it's the capabilities represented by the staff in creative approaches, as well as commitment to vertical data integration that will ensure success.

Network Approach

In this approach, the objective was to explore vertical data integration through a regional agency with an existing operational geo-based information system. The Association of Bay Area Governments (ABAG) was to integrate Landsat data into an existing environmental data base and then assess the usefulness of Landsat data to ABAG and other Bay Area users.

The specific project ABAG chose for this approach was to provide input to existing air pollution models in the form of information regarding the location and extent of certain vegetation types. There is evidence that vegetation, under certain conditions, emits significant amounts of hydrocarbons and this was a way to assess the total amount for a large area. The Bay Area Spatial Information System, ABAG's existing GIS, contained environmental data covering the San Francisco Bay Region but not a consistent vegetation cover map. The CDF data set was again to be used for the vegetation component and after modification to a suitable classification would be input to BASIS. Finally, the vegetation data, combined with estimates of hydrocarbon emission rates for the different vegetation classes, was used to produce models and maps identifying the amount and location of emissions.

Two major technical problems were encountered in trying to use the CDF data set: the vegetation classification was too general for this specific application and the positional location of the data was not accurate enough for inclusion to the ABAG data base. Representatives of both ABAG and NASA working on systems at NASA Ames were able to correct both these problems without much difficulty, however. As an approach to vertical data integration, this networking model was technically successful primarily due to the fact that data was being transferred to an existing GIS with well-defined parameters for operations. That ABAG did not have Landsat processing capabilities was not at all limiting since Ames was close by and, for future work, there is no technical barrier to ABAG acquiring that capability.

As a result of this project, there exists a San Francisco Bay Region vegetation classification in the BASIS data base that could be accessed by the various users of ABAG's services. The extent to which it might be used further depends on ABAG (1) informing others of its existence, (2) helping them work with it, and, importantly, (3) keeping it up to date.

Developmental Approach

Monitoring prime agricultural lands was the major focus of this approach which was conducted by the Geography Remote Sensing Unit of the University of California at Santa Barbara. The objective was to work with users at the county level (Ventura County and East Fresno County) to map land-use change using Landsat and update prime agricultural land maps. The intent was to foster cooperation between different levels of government (county, state and federal).

The conversion of prime agricultural lands to urban land use is a very emotional issue within the State of California. Due to the vast area covered, the speed of conversion and frequency required to update statistics concerning the problem; it also is a difficult problem to track. This is compounded by the fact that there are actually three means by which prime agricultural lands are converted and Landsat is best suited to monitor one of them (urbanization) and not the other two (soil erosion reducing the prime classification and economic factors that take the land out of production). UCSB endeavored to identify prime agricultural lands from Landsat first with single-date classification and then multirate classification. It was determined, however, that this couldn't be done accurately enough from Landsat alone. Current efforts are directed at multirate Landsat classification to detect land-use change which can then be identified by more detailed examination. This latter approach, developmental in itself, is technically very promising. It puts emphasis at the state level for detection and county level for identification. It remains to be seen if this will lead to an operational system, but the outcome will most likely be governed by institutional factors than technical.

Within this entire approach was a specific technical problem UCSB was particularly interested in solving: how to aggregate detailed county-level information to report it at

state-level required resolution. Choosing prime agricultural lands as a problem area happened to contain so many institutional issues that this aggregation issue was never addressed fully and remains as a technical issue to be considered.

Industry-Assisted Approach

The objective of this approach was to integrate Landsat data (again, the CDF data file) into a local operational data base with private industry to provide the GIS capability. The vertical data integration aspects were to come from the cooperation and data transfer among county and federal governments and a public utility. The concept was to build a data base covering the County of San Bernardino using inputs from the County Planning Department, the U.S. Forest Service and Southern California Edison and then exploit that data base to determine how useful the Landsat data might be for monitoring growth within the county. As the project proceeded, more slowly than expected due to contractual issues, the data base exploitation objectives expanded to include investigation of additional potential uses for the integrated data base. This was to be essentially a two phase project with data base construction first, followed by data base exploitation; consequently, there are not many final products to evaluate at this time. The technical issues encountered to date are concerned with data transfer, specifically Landsat data being entered to the county data base.

Representatives of NASA and Environmental Systems Research Institute, the private industry contractor, worked on computer systems at NASA Ames to modify the CDF classification to meet the project objectives. As in the ABAG project, the CDF classes were too general for the county land use mapping requirements. Highly knowledgeable people were required to perform extensive post-classification stratification and then reclassification within strata to achieve an acceptable land use classification. In addition to the thematic limitations, the CDF classification

ORIGINAL PAGE 13
OF POOR QUALITY

was not as accurate or precise in terms of geographic location as was needed to register to the other layers in the data base. However, this problem was also solved to finally allow the integration of the CDF data set into the county base. To monitor growth, a multirate Landsat data set was next considered--the CDF data represented conditions in 1976 and an additional data set for 1979 was to be added to reflect, within the data base, areas that had changed during the three years. This change detection "mask" was eventually created after examining three alternative approaches.

Technically, this approach had a similar environment as the network approach in that there was an existing GIS to which the Landsat data sets were transferred. NASA representatives again provided the Landsat processing expertise necessary to modify the CDF data to suit the requirements of this application. The success of this approach, however, remains to be seen. From a technical standpoint, the issue is if the county and ESRI representatives can satisfy any Landsat processing requirements for future work in terms of both systems and knowledge. Institutionally, the issues are county commitment to continuation and costs of the products the technology is providing. The technical issues can most likely be solved, and while the institutional ones are still uncertain, the concept of private industry to supply and maintain data bases for public agencies on a contract basis is worth further consideration given the high cost of developing any in-house capability.

In summary, there were no technical problems encountered in any of the approaches that could not be solved. In other words, there were no purely technical barriers to vertical data integration in any of the approaches. As for development of any operational systems, it would be premature to draw any conclusions concerning using any one of these approaches as a model. While each approach exhibited positive results, institutional issues still are an overriding concern. Given specific target

ORIGINAL PAGE IS
OF POOR QUALITY

applications, commitment and involvement on the part of the users and access to appropriate, affordable hardware/software systems (either in-house or in the private sector) are critical to achieving vertical data integration. The final section of Paul Wilson's Institutional Issues report contains ideas on how to structure work towards specifying or selecting a method of implementing operational systems in California. Given the apparent lack of difficulty in handling technical problems noted in the four approaches to date, the steps described in that section represent the major keys to achieving an operational vertical data integration capability within the state.

ORIGINAL PAGE IS
OF POOR QUALITY

**OPERATIONAL ALTERNATIVES FOR LANDSAT IN CALIFORNIA:
INSTITUTIONAL ISSUES**

**PAUL M. WILSON
GEOGROUP CORPORATION
BERKELEY, CALIFORNIA**

CONTENTS

<u>INTRODUCTION</u>	1
LANDSAT AND NATURAL RESOURCES DATA	1
THE CIRSS PROJECT	2
VERTICAL DATA INTEGRATION	2
ORGANIZATION AND SCOPE OF THE REPORT	4
<u>IDENTIFICATION OF OPERATIONAL ALTERNATIVES</u>	6
ESTABLISHING A DEFINITION	6
FOUR APPROACHES	7
EVALUATION CRITERIA	8
<u>EVALUATION OF THE FOUR APPROACHES</u>	9
METHODOLOGY	9
DEVELOPMENTAL : UCSB	10
NETWORKING : ABAG	12
EVOLUTIONARY : CDF	14
INDUSTRY ASSISTED : ESRI	16
<u>CONCLUSIONS</u>	18
COMPARING RESULTS	18
NEXT STEPS	20
<u>REFERENCES</u>	22
GENERAL REFERENCES	22
PROJECT DOCUMENTATION	23
<u>APPENDIX : EVALUATION CRITERIA</u>	25

INTRODUCTION

LANDSAT AND NATURAL RESOURCES DATA

The growing attention to environmental issues has rapidly increased the demand for information about natural resources such as geology, soils, and vegetation. Information about environmental hazards, such as flooding and earthquakes, has also become more important. These types of data are most useful when they can be used in conjunction with similar information about the patterns of human habitation (land use, transportation networks, boundaries of political and administrative units, energy facilities) to study the consequences of mankind's interplay with the natural environment.

Most of this resources data, to be useful in making decisions about the utilization of resources or the reduction of hazards, must be closely linked to a known geographic area. It must be possible to compare the boundaries describing each data type to the boundaries of related data types. The location of earthquake faults, for example, might be compared with land use patterns to help assess potential damage, or information about vegetation may be used to reduce fire hazards for residential areas.

Applications such as these have created demands for an increasing amount of data, for methods to locate that data in some type of spatial framework, and for the analytical capability to overlay and compare several different data sets. These demands have outstripped manual methods of handling mapped information; the result has been the creation of many computer-based geographic information systems in both government and private organizations.

California, because of several factors - the complexity of its natural environment, the large size of most political jurisdictions, and comparatively strong laws dealing with resource issues - has been the site of many of these systems. (A 1978 report, Resource Information Systems in California Local Government, cataloged fifty active public sector systems. There are also several systems in use by state and federal agencies, as well as those operated by utilities and timber companies.)

Although many of these systems are being used successfully, there is an inconsistent spread of coverage over the state. Some areas are included in several overlapping systems, while others are not covered at all. There is little cooperation among these active systems; little sharing of data or of processing capability takes place, even when systems contain the same type of data for the same geographic area, and use it for the same type of applications. Each organization maintains its own data base, which often duplicates much of the contents of other data bases.

**ORIGINAL PAGE IS
OF POOR QUALITY**

This duplication can be overcome where there is a common source of data for all users. Landsat represents a source of this type. Landsat, along with similar remote sensing technologies, provides a data source with major advantages in coverage and detail for looking at resource issues. That potential application of Landsat is a fundamental theme of this report.

THE CIRSS PROJECT

These trends, the growing need for resources data in California and the potential of Landsat for meeting much of that need, resulted in the creation of a new program in 1979. This project, started by NASA and the state Environmental Data Center, is called the California Integrated Remote Sensing System (CIRSS).

The CIRSS program is guided by a Task Force, which is composed of organizations representing the entire range of Landsat users. Included are governmental agencies (local, regional, state, and federal), academic institutions (universities and colleges), public utilities, and private industry. The Task Force is responsible for selection and evaluation of projects as well as for overall policy direction.

This report, which is part of the operational alternatives study, is one element of the CIRSS program. Its relationship to the overall framework of the CIRSS effort is that of observer; the operational alternatives study has reviewed progress of the other projects over the past year, and this report is a summary of that review process.

VERTICAL DATA INTEGRATION

One of the issues described above, the duplication of data collection and data handling efforts, points to the problem of vertical data integration. Conceptually, there could be for any given geographic area a shared mechanism for collecting, storing, analyzing, and displaying resource data about that area. (Or alternatively, a common means for collecting and updating data even if it is stored and processed in different places.) This mechanism should be accessible to every public and private organization which makes planning decisions affecting the area.

The problem becomes more complicated because of the wide differences in area coverage by different organizations. A city government is normally interested only in data that describes its own jurisdiction; this area fits within that of a county government and a regional agency, and all fit within the geographic scope of state and federal agencies. Other organizations are interested in areas that do not conform to political

ORIGINAL PAGE IS
OF POOR QUALITY

boundaries, or may be studying a problem (such as air pollution) which includes parts of many areas. Further, the level of detail needed and other data characteristics will vary significantly among different users.

The issue, then, becomes a very difficult one to define precisely. The concept - a common resource data base, or at least the transfer of a single compatible data set to several data bases - is known as vertical data integration. It has been a major focus of the CIRSS program. The CIRSS Task Force, after much discussion, adopted the following definition:

Vertical Data Integration refers to the general compatibility of data formats, classification methods, and encoding routines whereby data collected within a geographical area by one level of government and its associated agencies can be selectively incorporated into the geo-based information systems of many other levels of government and their associated agencies with minimal data manipulation or reformatting.

Obviously, data integration is not something that happens automatically. It can be achieved only through deliberate and concerted effort of all involved organizations.

In 1979 and 1980, the CIRSS program concentrated on identifying the conceptual issues around vertical integration. One result of this effort was a report, Elements of Vertical Data Integration. This report discussed a group of technical and institutional factors which are necessary to achieve integration:

TECHNICAL ELEMENTS

1. DATA CLASSIFICATION
2. DATA STATUS
3. DATA ACCURACY
4. DATA VOLUME
5. TEMPORAL FACTORS
6. GEOGRAPHIC COVERAGE
7. DATA REPRESENTATION
8. RESOLUTION
9. GEOGRAPHIC REFERENCING
10. POSITIONAL ACCURACY
11. DATA TRANSFER MEDIUM

INSTITUTIONAL ELEMENTS

1. DIFFERENCES IN NEEDS
2. INTERPRETATIONS OF DATA
3. LEAD TIME
4. USER AWARENESS
5. ORGANIZATIONAL RESPONSIBILITY
6. COST FACTORS

ORIGINAL PAGE IS
OF POOR QUALITY

7. STAFFING AND EQUIPMENT
8. DOCUMENTATION
9. LEGAL BARRIERS
10. PUBLIC / PRIVATE ISSUES
11. POLITICAL PROBLEMS
12. PERSONALITY FACTORS

The current work follows closely on this framework, but moves from the conceptual definition of data integration into the actual process of implementing it. This work falls under the heading of operational alternatives, and involves the identification and evaluation of alternative approaches for developing and implementing statewide systems for handling Landsat and other resource data in an integrated fashion.

ORGANIZATION AND SCOPE OF THE REPORT

This report has five main sections. This introductory chapter is designed to cover the background - to describe the problem to be solved and to provide a setting for the issues to be discussed in detail later in the paper.

Definitions are the focus of the second section. Here we will attempt to reach a more precise definition of the term operational alternatives, and to describe the CIRSS process of building operational systems in California.

The third section, which makes up the bulk of the report, looks in detail at each of the four approaches and the project used to test it. A brief project description starts the discussion of each project, and an evaluation makes up the rest of the subsection.

This analysis of the four approaches separately provides the raw material for the next section, which compares the approaches and looks for general findings about their progress to date. Recommendations for the next series of actions to implement operational systems are also covered.

Much written and oral documentation has been produced by the CIRSS program so far. A final section on references contains a guide to this material, and also lists some useful publications of a more general nature.

Much of the scope of this report has been covered implicitly above in describing the meanings of vertical data integration and operational alternatives. It should be clear that we are dealing with natural resource data (particularly Landsat as a data source), and not with other types of information systems. Consequently, all of this data must

ORIGINAL PAGE IS
OF POOR QUALITY

be spatial in nature, and geographic location is a key component in its description. We are not, then, concerned with the implementation of systems for processing non-spatial data such as budgets or personnel records, even though the organizations involved must deal with these issues also.

Some attention has been paid to manual map overlay systems, but the emphasis is on automated methods. This bias, which coincides with a strong overall trend in the handling of geographic data, is reinforced by the digital nature of Landsat imagery.

The State of California is the geographic focus. Although this study has reviewed the experience of existing operational systems in other states and may have some potential value for other areas considering natural resource systems, the emphasis is on establishing systems for California. The primary goal is a statewide system, although this idea does not preclude support of substate systems where they are compatible with statewide goals.

Since this effort is aimed at the creation of operational systems, it must deal with the economic and political realities which exist in California. It is not an academic or theoretical study of possible organizational structures, but is limited to those which have a real possibility for implementation.

Finally, this report should be seen in the perspective of past and future CIRSS efforts. It is based upon previous Task Force direction, and aims at producing information which can guide future efforts.

IDENTIFICATION OF OPERATIONAL ALTERNATIVES

ESTABLISHING A DEFINITION

The meaning of the term operational alternatives as used in this project is fairly well understood, and can be expressed (as in the CIRSS Program summary) in few words: "evaluating alternative approaches for developing and implementing a state Landsat data distribution network." There is a need, however, to clarify this description and to explore it in more specific terms.

By operational we mean, of course, something that is operating or functioning. The implication is that to be operational, a system would be capable of day-to-day operations over a long period of time. In the context of the CIRSS program, it is also understood that the system would be operated by (or at least managed by in a policy sense) some agency of state government, and would not count on receiving continued technical or financial support from NASA.

The definition given above is a bit misleading in one sense: it emphasizes Landsat data to the exclusion of other remote sensing data sources or other types of resource data. Landsat is clearly a major emphasis, since the CIRSS program was started through support from NASA and since Landsat does represent a data source with great potential for exploring resource issues. It is important to note, however, that the concept of operational alternatives as seen by the Task Force extends to other types of resource data and to the geographic information systems required to handle this variety of data.

Some discussion about the nature of the target system is also necessary. As mentioned above, vertical integration may range from a complete integration of systems (with all data collection and processing for many users taking place on one system) to simply data integration (where a data set collected through a common effort was processed on several separate systems). The CIRSS concept is to support all types of integration, with a long-range goal of achieving some level of systems integration so that users without their own processing capabilities are not left out. Furthermore, it is assumed that many systems will be needed in a state as large and diverse as California; there is no bias towards the creation of a single monolithic statewide system.

ORIGINAL PAGE IS
OF POOR QUALITY

FOUR APPROACHES

The other part of the term is alternatives. Since the procedures for establishing an operational system in California are by no means obvious, the CIRSS Task Force identified four alternative approaches to be tested. These approaches were identified by looking at how other states had established operational systems, and by assessing the current status of Landsat processing systems in California (agencies using remotely sensed data, agencies having GIS capability, and applications which would be a good test of Landsat as a data source).

A common technical base was established by defining the use of an existing data set, the statewide Landsat classification done in 1979 for the California Department of Forestry. Each approach was to focus on how this data set could be integrated into a specific application (or a related set of applications) in a specific organizational setting.

After consideration of many scenarios, it was decided to test the following four approaches:

*Sounds like
actual project
rather than
initial approach* [

Developmental - Transfer of an existing methodological base, developed at a university, to several governmental agencies. The GIS framework in the agencies was in a developmental stage, as was the application of remote sensing technology to the chosen problem.

Networking - Integration of Landsat data into an existing GIS system operating at a substate level. This is a distributed processing approach: creation of a statewide system through the networking of several regional systems within the state.

Evolutionary - Continuation of applying the original CDF data set within the same organization. Using the same data, but looking at how the agency's use of it evolved, how other data sets were combined with it, and whether other organizations would find it applicable to their own needs.

Industry Assisted - Utilization of system design and processing capability within the private sector. Government agencies as end users, but without internal processing capabilities.

A specific project was then chosen to represent each of these approaches. The discussion of these projects, as they reflect findings about the generic approaches, is the major part of this report.

EVALUATION CRITERIA

After the four approaches had been defined and projects to test them had been chosen, it was necessary to create a method for assessing the results. This method took the form of a set of evaluation criteria, discussed and approved by the Task Force. They are listed in the Appendix to this report. These criteria, which were framed as a series of questions, were designed to provide a standard way of assessing the progress of each project. They were aimed at providing information for the Task Force to evaluate each approach - not just the project which represented that approach. As a result, the criteria seek to identify general findings and do not seek a narrow view of the project's conduct.

Some of the criteria are clearly more important than others in this evaluation. Questions which relate to the specific project, rather than to the approach, are necessary but do not yield the information about operational alternatives which is the real objective of this process.

EVALUATION OF THE FOUR APPROACHES

METHODOLOGY

Evaluation of operational alternatives should be viewed as a continuing process, broader in time and in scope than can be shown in a written report. This evaluation process represents close monitoring of the overall CIRSS program as well as the individual projects. It has been conducted over the past year through participation in Task Force meetings and periodic discussions with project leaders. The evaluation process will continue beyond this report, as further work on the projects reveals new insights.

It should be emphasized that this is an evaluation of approaches, not of the projects themselves. The CIRSS objective is to use the projects as a way of testing the alternative approaches, so here we concentrate on those findings which can be generalized to the creation of operational systems. The emphasis is on results that reflect on a general approach rather than events which derive from the particular circumstances of the test project.

Three types of sources were used extensively. The first consisted of materials developed for and by the Task Force during the past year; these yield a picture of the progress on each project over time. The second source is the written documentation produced by project staff and NASA personnel. This material contains detailed background information about each project and the problems it has encountered. A third source is a series of interviews with project personnel (users, agency staff, NASA and contract staff) conducted during January, February, and March of 1981.

The CIRSS evaluation criteria, which were discussed in a previous section, were used as a foundation for this work. They were analyzed to determine the most meaningful issues (in the context of the current setting in California), and then provided the structure for interviews and review of the written project documentation.

Although this report on institutional issues and the companion effort on technical issues were completed under separate contracts, there has been a close working relationship in preparing them. Considerable assistance was also provided by CIRSS project staff and NASA personnel.

This evaluation is not a strict scientific test of the approaches. Since it is designed to look at how the alternatives fare in real situations, it has been subject to complications which may not exist in a real operational setting. For example, the split of funding and responsibility (money from NASA, direction from a user agency) was cited as a problem in some projects. In an actual operational situation, funding and direction would normally come from the same agency.

DEVELOPMENTAL : UCSB

PROJECT DESCRIPTION

The Geography Remote Sensing Unit of the University of California at Santa Barbara carried out the project which represented the developmental approach. The project had as its focus the problem of change detection in prime agricultural lands. The primary end users were two local governments, with some involvement by other public agencies, such as the California Department of Conservation.

The conversion of prime agricultural land (PAL) to urban uses has been recognized as a major problem, but it has been difficult to develop programs to solve it. Even assessing the magnitude of conversion has been a serious problem in California, given the vast areas to be covered and the speed of urban growth in the state.

Landsat, with its frequency of coverage and ability to readily pick up detailed vegetation patterns, has obvious potential as a data source. This is particularly true where automated change detection techniques can be applied. The PAL issue, and the use of change detection techniques to study it, is the focus of this project. It is viewed as a developmental approach in that the tools to study the problem are relatively new, and a methodology must be developed as part of the process.

Two types of potential users, each of which plays a role in the PAL issue, were chosen. The first is a statewide focus, with the major role being taken by the Department of Conservation. The second is local government, with two counties (Ventura and Fresno) taking the local perspective and acting as test sites. The project itself involved identification of changed areas from multi-date Landsat analysis, along with the integration of this data into local geographic data handling systems.

In addition to the concept of a developmental approach, this project also looks at the state's institutions of higher learning as a source of assistance in implementing operational systems.

Evaluation

Results of this project seem, from a technical standpoint, promising. Landsat does indeed represent a data source which can detect the magnitude and location of converted lands with an acceptable accuracy level. The institutional results are less clear.

During the two years of activity on this project, the Department of Conservation has moved into a more more active role in the PAL

ORIGINAL PAGE IS
OF POOR QUALITY

preservation arena. This meant that the county governments, which were the original institutional focus, were not the major force in the field at the time of this evaluation. This experience points to a frequent problem in long-range projects: agency roles can sometimes change dramatically between the project design phase and actual implementation of a system.

The present situation in the PAL area is, of course, a perfect test of vertical data integration. Can one data source, Landsat, provide information usable at both county and statewide levels? The results so far indicate a positive answer to this question. It appears that the process developed here can be used to update maps at different scales as needed by several users.

Communication between the Task Force and the project staff from UCSB was often poor. The Task Force was much more attuned to the shifts in political responsibilities, while the research staff from the university concentrated more on the technical issues.

Participation from the user agencies was weak at first; this aspect seemed to improve as the project progressed. This is an institutional issue that represents a potential problem in all university-related projects. Considerable effort must be made to direct the project at current user needs rather than at interesting research topics; a balance between technical and institutional efforts must be achieved.

Despite these potential drawbacks, it is important to remember that the universities represent a viable source of assistance for local governments. It is doubtful that either of the two test counties would have developed Landsat processing capability on their own - and they are some of the more sophisticated counties in the state in terms of geographic data use. University assistance, even with all of its potential problems, may be the only source of technical support in many instances.

NETWORKING : ABAG

PROJECT DESCRIPTION

The assessment of air pollution caused by vegetation was the focus of the networking project. Although it has been known that certain types of vegetative cover can emit significant amounts of hydrocarbons under certain conditions, there had been no way to assess the total amount of pollution from these sources for a large area (or to compare this amount to the pollution caused by manmade sources). The lack of good information about the amount and location of different vegetation types was a major problem; Landsat represented an ideal data source for this type of application.

The lead agency here was the Association of Bay Area Governments, with secondary roles played by the Bay Area Air Quality Management District, the California Air Resources Board, and the federal Department of Energy (through Lawrence Livermore Lab). Technical processing of Landsat data was performed by NASA.

The networking concept was embodied by ABAG's existing GIS, the Bay Area Spatial Information System (BASIS). This system was seen as a focal point for data about the San Francisco Bay Region. Landsat data could be added to BASIS as another level of data, and models to calculate emissions could be readily produced (and the results transferred as inputs to ongoing air pollution models). The model of operational systems in this alternative would be a collection of substate systems, networked by an actual or administrative communications system.

Three major steps were involved. The first was obtaining a suitable vegetation classification using the CDF file. Next, this file was to be integrated into the BASIS data base in the same form as its other data sets. The final step was to use this vegetation data, along with estimates of the emission rates for different vegetation classes, to produce models and maps identifying the location and quantities of hydrocarbon emissions.

EVALUATION

All of the short-term technical and institutional objectives of this project were achieved successfully; it is uncertain at this point whether some of the longer range goals will be met.

The chosen application was an excellent example of vertical integration, both in the data integration sense (using the CDF Landsat file as a base) and in the system integration sense (several agencies, at different levels of government, using BASIS as a common processing capability). This concentration of the technical work at two places

ORIGINAL PAGE IS
OF POOR QUALITY

(NASA for the Landsat processing and ABAG for the GIS work and model output) minimized the problems which may have resulted if each of the participants in this type of project had done their own technical processing.

Early and explicit identification of the application was a major advantage in this project. The air pollution issue is clearly one of great importance, and one which had been studied extensively. This project was narrowly focused on one aspect of that issue; the vegetation contribution to pollution had been identified in earlier studies as an unresolved problem, and there was broad support for trying this new tool to assess it. The project was also clearly focused in a time sense: it was seen as a discrete product to be delivered at a specified date, with long range considerations flowing from that first product.

Another advantage derived from the mix of organizations which participated in the project: each of the agencies involved has a distinct mandated role in the overall air pollution planning process. This gave a strong institutional base to the project. Although there was some initial misunderstanding among the users about Landsat technology, this potential problem was quickly overcome. This experience points to the importance of early training programs for potential users.

Existence of an operational GIS, BASIS, was a positive factor in both technical and institutional senses. The technical advantage of an existing GIS meant that integration of the Landsat data and subsequent modeling were comparatively easy. The institutional factor is less obvious, but equally important: ABAG had staff used to dealing with spatial data in an automated framework, and Landsat could be viewed as just another data set in the regional data base.

The primary short-term goals of this project were completed with relatively few problems. Since this project (in contrast to the other approaches) was seen in the CIRSS program as a one-year effort, its results are comparatively easy to evaluate. It can almost be viewed as a final assessment rather than a project report. The results are, overall, very promising.

Two of the long-range goals remain unmet, however. These are the creation of an operational Landsat processing system at ABAG (using ELAS) and the formation of a GIS Council to guide geoprocessing policy in the Bay Area. Uncertainties about future funding, as well as the departure of a key staff person, have limited ABAG's ability to support new system capabilities. Although the single Landsat-derived land cover file is now a permanent part of the BASIS data base (and has already been used for other applications), there is still no operational way to process other Landsat data for the region. ABAG staff have gained a considerable amount of training and capability in use of Landsat, but software transfer is necessary for there to be a true operational system.

ORIGINAL PAGE IS
OF POOR QUALITY

EVOLUTIONARY : CDF

PROJECT DESCRIPTION

The technical base for testing the four approaches was a Landsat file processed in 1979 for use in forest resource assessment by the California Department of Forestry. In this approach, the evolution of that file and its applications is monitored. The basic assumption in this alternative is that the CDF capability (staff expertise as well as the existing Landsat file) would evolve to a point where it could help support local agency vertical integration projects.

Other applications have been made in other CDF programs, in U.S. Forest Service studies, and in several county projects. Of particular importance here is the creation of a detailed GIS for Santa Cruz County, with participation by the local government as well as assistance from several state and federal agencies.

Much of the work on this project during the past year has concentrated on defining the hardware and software capabilities needed to create an operational system in Sacramento. Such a system would be the focal point for further development of CDF capability as well as possible assistance to local governments.

EVALUATION

The CDF project, seen in its full scope of several years effort, is perhaps the most ambitious of the four CIRSS tests. It is larger in scale, involving both a larger geographic area (the entire state in some of its components) and a broader mixture of participants.

This very scale has presented some problems. The definition of applications has been diffuse, largely because of a continuing tension between planning and operational functions at CDF. A single policy direction has been difficult to maintain over the life of the project; change is, after all, a deliberate element in this approach.

Some institutional problems have developed during the process so far. These derive primarily from the differing perceptions of NASA and the user agency. For example, the initial statewide classification was successful, but (from the viewpoint of CDF) too heavily emphasized the product rather than the process. Also, NASA's vast computing resources often make it difficult for the agency's technical staff to appreciate the funding limitations under which a state agency must operate.

The products derived from this project are potentially useful in substate (vertical integration) applications, but the test of CDF's ability to run them in an operational setting has not yet been

ORIGINAL PAGE IS
OF POOR QUALITY

completed.

The GIS development in Santa Cruz County is the clearest test of vertical data integration in this project. Although it was not fully implemented at the time of this evaluation, this work looks very promising as an example of data integration concepts.

A "distributed analysis" model has been one interesting institutional idea to evolve out of this project. This method divides Landsat processing and analysis among universities, with some done centrally by state staff. A possible advantage of this approach is the use of people more familiar with each geographic area covered in a specific application.

INDUSTRY ASSISTED : ESRI

PROJECT DESCRIPTION

Integration of Landsat data into multipurpose local data bases was the focus of the fourth project. ESRI, the Environmental Systems Research Institute, was the contractor; the San Bernardino County Planning Department (which must cope with the management of urbanization in a very large area with severe growth pressures) and the U.S. Forest Service (which must handle fire problems made more critical by that urban growth) were the end users. A large utility, Southern California Edison, has also participated to some extent.

This project looks at the provision of data base capability through a private sector firm, ESRI, to users who may have a diverse set of applications. It is similar to the UCSB project in that the Landsat processing and GIS capabilities may not be institutionalized by one of the user agencies, but may remain in an external organization which would provide services as needed. (NASA personnel actually provided the Landsat processing work on this project; there was extensive training and transfer of capability to the industry representative, however.)

Since this is a two-year project, there are few final products on which to base an evaluation. Work during the first year has concentrated on assembling a data base, and major applications will be performed during the second year of the project.

Some of the tasks done to date, such as creation of a data set showing land cover change from 1976 to 1979, appear to show applicability to many types of land use studies. Also, the assembly of a common data base from multiple sources is an interesting exercise in both technical and institutional skills.

EVALUATION

This project started slowly, largely because of contractual problems between ESRI and NASA. These problems were magnified by uncertainties about Task Force direction and the responsibilities for overall project management. Although these particular difficulties may be specific to the project and do not necessarily present a real problem in the industry-assisted approach, they do point to the differences which must be taken into account when a third party is involved.

Another factor which may be limited to this test setting, and not applicable to an operational system, is the split between funding and responsibility. There was considerable confusion about responsibility for overall direction and specific decisions, since NASA was providing the money and other agencies were the designated users. In a normal

ORIGINAL PAGE IS
OF POOR QUALITY

industry-assisted project, the user agency would also be providing the funds and would therefore be more highly motivated to follow up its investment.

Lack of specifically-defined products has presented some problems throughout this project. Both the County and USFS started with a list of general application areas in which they were interested, but there were no specific products identified until very recently. As a result, the data base building process has taken place without a clear statement of priorities among user needs. This has not been a severe problem, however, since ESRI (because of its experience with GIS applications in general and this geographic area in particular) seems to have anticipated needs well enough so that the expected uses during the second year can be accommodated. The potential for a problem here should not be ignored; a private firm with less expertise in a specific area may have not been as successful.

Also, this lack of a specific product to drive the project has resulted in a subtle institutional problem. The technical process in this type of application must be continually driven by recognized organizational needs or there is a tendency to forget application objectives and concentrate on other parts of the process. This project did not establish an early link between applications and information needs, so much of the effort has gone into data base assembly. One result has been low participation by user agencies.

Monitoring the project uncovered several other institutional factors. In this case, the private firm had a great deal of GIS background, but little direct experience in Landsat processing. A firm with different experience (for example, heavy Landsat processing but little GIS use) may have handled the initial technical work more easily, but would have had much more trouble in other aspects of the work.

The value of Landsat data in a GIS was shown in the example of a planned USFS application, establishing a greenbelt for fire damage prevention. This idea was largely motivated by a disastrous fire in the study area. The inherent flexibility of a GIS can be an important institutional factor in attracting real applications.

Much of the success of the industry-assisted approach depends upon costs. It is still too early in this project to assess costs of particular products, but the user agencies have shown some concern over being tied to a private firm for all future processing. Public agencies, given the current fiscal situation, often have no choice; they cannot afford to build internal capability, and must seek technical capability from private firms if the project is to be done at all.

A final institutional factor is the ability of private industry to serve as a facilitator for data transfer. The use of utility data in the county application was made much easier because of the private firm's knowledge of both organizations. This example may point to one of the major advantages of the industry-assisted approach.

CONCLUSIONS

COMPARING RESULTS

It should be clear from the above descriptions that there were significant differences among the projects in how successfully institutional barriers to vertical integration were overcome. The ultimate objective of the operational alternatives project is the synthesis of those differences into a meaningful and coordinated set of findings which can be used as a foundation for further work.

There are several limitations which must be recognized in making comparisons between projects. Each of the projects was different from the others in many ways - not just in the approach it represented. Since start times and overall time frames were quite different, the projects were at different stages of completion when this report was done. This study, then, is a look at project status at one point in time; it is more of a progress report than an assessment of final results. It should also be recognized that the projects differ somewhat in the scope of their work, and this may be reflected in apparent success or failure.

Some extraneous factors are difficult to filter out. For example, some of the project participants have also played major roles on the Task Force, which should have added a closer adherence to objectives than might have been possible with the other projects. Also, the author of this report participated in one of the technical tasks in the ABAG project, so additional knowledge about that work gives it a different perspective from the other projects.

Despite these limitations, there are still a number of valuable things to be learned from the CIRSS demonstration projects so far. None of the approaches has been proven to be incapable of producing an operational system; on the other hand, none demonstrate a guarantee of success. Each, based on current results, shows potential and must still be viewed as a candidate for implementation.

The evaluations done for this report have uncovered a number of general principles which appear to be useful in designing operational systems.

Definition of a target application seems to be a major factor. Clear and early specification of deliverable products helps gain organizational support (as well as having obvious technical advantages). The UCSB and ABAG projects seemed to benefit greatly from having a clearly defined set of products, while the other two projects started with a more general list of potential application areas.

The definition of products aligns closely with breadth of scope ⁽⁻⁾ magnitude of the products, timing, and number of participants. The ABAG project, with a single major product and a relatively short time frame, had much less of a problem in goal definition than the other projects. There were fewer institutional variables to consider. And in a shorter period of time, the institutional setting was less likely to change (as with the personnel turnover problems in the CDF project, and a change in organizational roles in the PAL issue).

Number of participants is a thornier issue. Having fewer organizations involved naturally limits some coordination problems; the CDF project has, on the other hand, been burdened by a growing number of participants and a corresponding difficulty in arriving at consensus goals and priorities. The whole thrust of data integration, however, is the inclusion of many agencies with common data needs. Much of the CDF coordination problem has resulted from a deliberate restructuring of the project to meet this goal.

Does this mean that vertical integration is more trouble than it is worth? The success of the ABAG project in reaching its short-term objectives, even with participation by several levels of government, would indicate that data integration is still a reachable goal. It is important to choose participants which agree on (and need) a small number of products; the project cannot be viewed as an effort to build a system that will do everything for everybody.

Geographic distance appears to be a surprisingly important factor. Close proximity between all participants makes institutional coordination as well as technical cooperation much easier. Cuts in travel funds in several levels of government have caused problems in each project; these problems were minimized where single day travel by automobile or transit was possible. This factor favors a network approach, which would divide the state into more manageable geographic areas; regional agencies or local academic institutions could act as nodes for such a network.

The existence of a GIS (staff expertise and experience as well as hardware and software) accessible to the user agency is another important factor. Landsat is most powerful when viewed as one part of a comprehensive GIS. An agency with GIS capability is much more likely to use Landsat products when they can be viewed as an element of an ongoing system. Organizational familiarity with GIS capabilities and limitations means that less training is required.

It is important to remember that the GIS capability need only be accessible to the agency, not necessarily operated by it; industry or a university may actually maintain the system and make its products available to the governmental agency. In theory, it may be most efficient to perform the work in-house (where the end user does the technical work) but practical considerations often make this impossible.

What do these comments mean in comparing approaches? An overriding factor seems to be user commitment; there must be a strong motivation on the part of the user agency (or agencies) no matter who does the

ORIGINAL PAGE 13
OF POOR QUALITY

technical work. The project cannot be driven by an outside organization (a university, a private firm, or NASA) but must be closely tied to a real and continuing organizational need. The technical advantages of outside assistance must be coupled with internal direction.

Another major factor is selection of the application. It must be broad enough to meet the needs of several users (assuming a vertical integration goal) and yet achievable. Working with a relatively small geographic area or a smaller organization may play an important role here; in the developmental approach, for example, the Santa Cruz GIS effort is gathering more interest than most statewide applications because its scope is more easily grasped.

In summary, each approach has shown some positive results. Each has also revealed weaknesses. More work is needed to select a method for implementing operational systems in California; the next section contains ideas on how to structure that work.

NEXT STEPS

The CIRSS Task Force faces a critical set of issues in the coming year. Federal funding for NASA programs is being cut severely, and fiscal problems at the state and local level add to the problem. Unless significant progress in establishing operational systems is made this year, the opportunity may be gone.

Several key steps are important in closing in on the implementation issue. These steps should be viewed as general guidelines for Task Force action, seen from the perspective of the operational alternatives study.

Concentrate on GIS implementation. Projects which work with Landsat alone, and treat other natural resource data as ancillary sources, miss the real emphasis of most government applications. Most real projects require a variety of data types; Landsat is best seen as one layer in a multilevel data base. Unless the capabilities (both technical and institutional) of a GIS are present, the application has little chance of success. The emphasis, then, should be on building strong GIS systems as a foundation for Landsat use, or on adding Landsat capability to operational GIS sites.

Push for a mandate. There are several agency contenders for the role of natural resource data coordinator in California. No state has been able to implement and maintain an operational system without resolving this issue of choosing a single focal point. The Task Force, which represents a broad spectrum of data users and a wide range of technical expertise, should play an important advisory role in this decision.

Address basic needs. Some of the fundamental elements of an operational system are still unmet in California. The enormous volume of data in

the state overwhelms many potential users. While everyone strives to add more data to the pile, there is no effective way of finding out what now exists. The Task Force should explore support of designing such a system - a kind of data base about data bases, with the ability to select by subject, date, and geographic area. A system like this can be seen as a necessary foundation for the implementation of operational processing systems.

Cut fragmentation. As a result of the factors noted above, there is a very high level of fragmentation in statewide efforts. There are many users performing many projects, using many data sources and many geographic information systems. Vertical data integration is exceedingly difficult in such an environment. Unfortunately, some of the CIRSS work has added to this duplication by supporting multiple geoprocessing systems in the same project. The Task Force must be careful to maximize its effectiveness by trying to coordinate data collection and GIS efforts.

Building operational natural resource systems in California is not an easy task. The state's size and complexity make any such undertaking an enormous challenge. The current high level of geoprocessing activity further justifies the need for operational systems, yet makes implementation of a common system more difficult. The CIRSS Task Force is in a unique position to identify and help implement an operational system. The opportunity for doing so in the coming year must be taken; it is not likely to happen again.

REFERENCES

This report is based on analysis of a large number of sources - reports, memoranda, notes on meetings, interviews. This concluding chapter lists the most important of those sources. It should serve as a guide for further work on the CIRSS program.

The references are divided into two major sections. The first contains general reference materials on Landsat and geographic information systems, and on the California situation and the CIRSS program. The second section covers each of the four projects covered in the report.

GENERAL REFERENCES

The following sources are useful for a general discussion of both technical and institutional factors important to geobased information systems.

Center for Geographic Analysis, University of Wisconsin. Data Needs and Data Gathering for Areas of Critical Environmental Concern. (Madison, 1975.)

National Conference of State Legislatures. A Legislator's Guide to Landsat. (Denver, 1978); A Legislator's Guide to Natural Resource Information Systems. (Denver, 1980); State Institutional and Technical Approaches to Landsat Utilization. (Denver, 1979.)

Natural Resource and Environment Task Force of the Intergovernmental Science, Engineering, and Technology Advisory Panel. State and Local Government Perspectives on a Landsat Information System. (Washington: Executive Office of the President, 1978.)

Michael Kennedy and Charles Guinn. "Avoiding System Failure: Approaches to Integrity and Utility." in Technical Supporting Report E of Information / Data Handling: A Guidebook for Development of State Programs. (Washington: Department of Interior, 1975.)

ORIGINAL PAGE IS
OF POOR QUALITY

Sources of information about the California situation include publications of the Environmental Data Center and the CIRSS Task Force.

Task Force minutes, 1979-1981

Annual CIRSS work programs

Viewgraphs prepared for NASA reviews

Paul M. Wilson. Resource Information Systems in California Local Government. (Sacramento: Office of Planning and Research, 1978.)

Paul M. Wilson. Elements of Vertical Data Integration. (Prepared for CIRSS Task Force, 1980)

PROJECT DOCUMENTATION

Developmental: UCSB

John Estes, Larry Tinney, Doug Stow. Conceptual Design and Preliminary Demonstration of a Prime Agricultural Land Component of an Integrated Remote Sensing System. (UCSB: Geography Remote Sensing Unit, December 1979)

John Estes (Principal Investigator). Prime Agricultural Land Monitoring and Assessment Component of the California Integrated Remote Sensing System. (UCSB: Geography Remote Sensing Unit, March 1980)

Tod Streich, Technical Presentation Outline on VICAR/EMIS Conversion

Interviews?

Networking: ABAG

Roberta Moreland and Gene Fosnight. The Development of a Biogenic Hydrocarbons Emissions Inventory for the San Francisco Bay Area, Work Plan and Working Papers. (Berkeley: ABAG, 1980 and 1981)

Presentation viewgraphs

Interviews: Roberta Moreland, ABAG Jan 21, 1981
Gene Fosnight, TGS Jan 21, 1981
Don Hunsaker, ABAG Jan 22, 1981
Don Olmstead, ABAG Feb 7, 1981

Evolutionary : CDF

Technical Memos

Presentation viewgraphs

Interviews: Dave Peterson, NASA Feb 6,1981
Jim Brass, NASA Feb 6,1981
Keith Maw, TGS Feb 6,1981
Nancy Tosta-Miller, CDF Feb 11,1981

Industry-Assisted : ESRI

Technical Memos

Presentation Viewgraphs

Interviews: William Likens, NASA Feb 6,1981
Keith Maw, TGS Feb 6,1981
Kenneth Topping, SBC Mar 16,1981
Ron Matyas, SBC Mar 16,1981
Jerry Christenson, ESRI Mar 16,1981
Russ Michel, ESRI Mar 16,1981